LANGUAGE DOCUMENTATION & CONSERVATION

Vol. 3, No. 2 (December 2009), pp. 241-246
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4442

LEXUS

from Max Planck Institute for Psycholinguistics Nijmegen

Reviewed by KRISTINA KOTCHEVA, *Humboldt-Universität zu Berlin*

1. INTRODUCTION. LEXUS is a web-based tool for manipulation and analysis of lexical data. It allows the creation of on-line multimedia encyclopedic dictionaries or lexica[1] from scratch or by importing structured data from other databases.

According to its developers, the tool is targeted at two groups of users: linguists, anthropologists, or other scientists contributing to and exploiting resources for linguistic and cultural documentation of languages and speech communities; and language activists documenting their own language and culture and creating a freely available lexical resource to be used by speech-community members for teaching and learning the language, thus supporting revitalization efforts.

LEXUS is specifically designed as a tool for the documentation and presentation of lexical data from endangered languages in the framework of DoBeS [German *Dokumentation bedrohter Sprachen* 'Documentation of endangered languages', see http://www.mpi.nl/dobes/]. The most interesting feature of LEXUS is obtained by combining it with its sister tool ViCoS [*Visualizing Conceptual Space*, see http://www.lat-mpi.eu/tools/vicos/]. This complements the linguistic information in LEXUS with multimedia, thus visualizing the meaning of a word or a phrase in its conceptual and cultural context. ViCoS and LEXUS interface with each other and allow the creation of arbitrary relations between lexical entries or parts thereof. Users can thus complement the lexical space with a conceptual space, creating any type of informal ontology to describe a given domain of interest as long as its concepts have a linguistic realization in the language(s) studied. Furthermore, ViCoS allows the creation of relations between objects in and across lexica.

Over the last three years, only a few DoBeS projects, along with some projects from within the Max Planck Institute for Psycholinguistics (MPI) Nijmegen, have built up experience with LEXUS. The launching of a new user interface in May 2009 might facilitate wider use of these tools in the future.

I became acquainted with LEXUS only recently when I started working as a new team member of KSDP [Kola Saami Documentation Project, see http://www2.hu-berlin.de/ksdp/]. KSDP is creating LEXUS dictionaries of endangered Kola Saami languages, all of which at present, however, are at rather preliminary stages (cf. Rießler and Ringersma 2009).[2]

---

[1] According to the manual and general feature descriptions provided by the MPI Nijmegen, LEXUS is a "lexicon tool" because the multimedia linking is a central part of its functionality. Documentary linguists, on the other hand often use "dictionary" in a synonymic way for illustrated glossaries, even if they are presented on-line and include different kind of multimedia.

[2] I am grateful to Michael Rießler and Jacquelijn Ringersma for introducing me to LEXUS and for useful comments and suggestions for this review.

E-ISSN 1934-5275

I am working with LEXUS using the newest version of the Mozilla Firefox (version 3.0.8), which is also the recommended browser according to the manual. I have a Macintosh computer with OS X 10.5.6. Since LEXUS is an on-line tool, the computer's operating system should be irrelevant. Problems with the interface or with character input encountered by me, however, might not be experienced by users of other operating systems.

The present review aims to provide an overview of the functionality of LEXUS (but not of its sister tool ViCoS), including importing and exporting data. The setup and the maintenance of a new lexicon project are explained in brief. Finally, some problems encountered during my own work with the tool are described, along with some reflections about making use of LEXUS in cooperative fieldwork with language communities.

2. OVERVIEW. The LEXUS tool is available on-line. It requires a web browser enabled by a Java script enabled web browser, preferably Mozilla Firefox. In principle, the tool is free for anybody to use. However, to start new projects, a workspace area on the LEXUS server has to be created first. This can still be done only by personal registration with the responsible MPI staff member. The LEXUS developer team is working on an automatic registration function from the homepage, which should be implemented very soon (Jacquelijn Ringersma, pers. comm.).

All data are stored in XML format on-line at the LEXUS server at the MPI. Structured data can be imported to or exported from LEXUS.

The LEXUS website [http://www.lat-mpi.eu/tools/lexus/] provides the user with a manual available both on-line and as a PDF file (Ringersma and Rybka 2009a). Note, however, that some parts of the manual are incomplete, as stated in them, because the implementation of the respective functions is still in development. Among these features are data import from XML, data export, applying sort orders, and the linking of media stored at the MPI archive.

Beside the manuals, the website provides a downloadable one-page guide to the tool. Furthermore, users can discuss questions concerning LEXUS with other users and the developer team through the LEXUS forum.

3. USING LEXUS.

3.1. STARTING A NEW DICTIONARY/LEXICON. After one logs into LEXUS with the help of the user account and password, the workspace opens in the Lexical Workspace Editor mode. The two other modes of the program are the Schema and View Editor and the Lexical Entry Editor. Switching between the modes is done through the LEXUS menu in the top frame. Each mode has two frames, each organized in several tabs.

Several lexica can be created and stored at one work space. The lexica are all listed in the Lexical Workspace Editor mode (in the left frame in the My Lexica tab which opens after the login).

Since the functionalities of LEXUS are dependent on one another, the following general work flow has to be followed:

1. Creation of a new dictionary/lexicon (either from scratch or through data import)

2. Definition of the data structure of the dictionary/lexicon

3. Definition of sort orders

4. Definition of the list view

5. Definition of the lexical entry view

6. Addition of new or revision of old entries and/or media

Once the new lexicon project is created in the Lexical Workspace Editor, the lexicon structure has to be defined in the Schema and View Editor. The lexicon structure in LEXUS consists of two components: one for metadata on the lexicon (called *lexiconInformation*) and another for the structure of the lexical entry itself (called *lexicalEntry*). The structure of lexical entries in LEXUS includes two default components: *Form* and *Sense*, but users may further build on the structure of the lexicon, according to the specific structure of the documented language, or the linguistic theory used. In the right frame of the program each selected element from the lexicon structure schema in the left frame can be displayed and defined according to its name, a description, and a specification, whether or not the given element is mandatory in the lexical entry.

Several sort orders can be created and stored at the workspace level and applied to each individual data category in any of the lexica. Once a new sort order has been created in the Sort Order tab in the Workspace Editor mode, the set of characters to be sorted can be defined. Any Unicode symbol can be included in the set of characters, including predefined combined characters or any combination of characters.

After defining the data structure of the lexicon and the sort order of its entries, one should decide on the format for the view of the single entries in the word list (displayed in the Lexical Entry Editor mode). This is done by choosing the List View tab in the Schema and View Editor mode by dragging the chosen data categories into a formatting preview window and applying certain formatting properties such as font size, shape, etc.

The lexical entry view is defined in the tab with the same name in the Schema and View Editor mode as well. Here, the data categories are not dragged into the formatting preview window, but are chosen from the Insert Data Category box at the bottom of the window. Users may, however, encounter some problems by using the formatting preview window, since it may not display properly on a screen resolution lower than 1280 x 1024 pt.

Instead of using the formatting preview window, the LEXUS manual recommends importing an externally created layout in HTML format as a better option. The LEXUS developer team provides support for doing this or creating example files that can easily be adapted for your own lexicon with the help of simple HTML editors.

Figure 1 shows a screen shot illustrating how a lexical entry (Kildin Saami *ābb'r* 'rain') looks in the Lexical Entry Editor mode. The List view is found in the left frame. The right frame shows (the upper parts of) the Lexical entry view, including multimedia. Note that formatting of the Lexical entry view is done with the help of an externally created layout in HTML format.

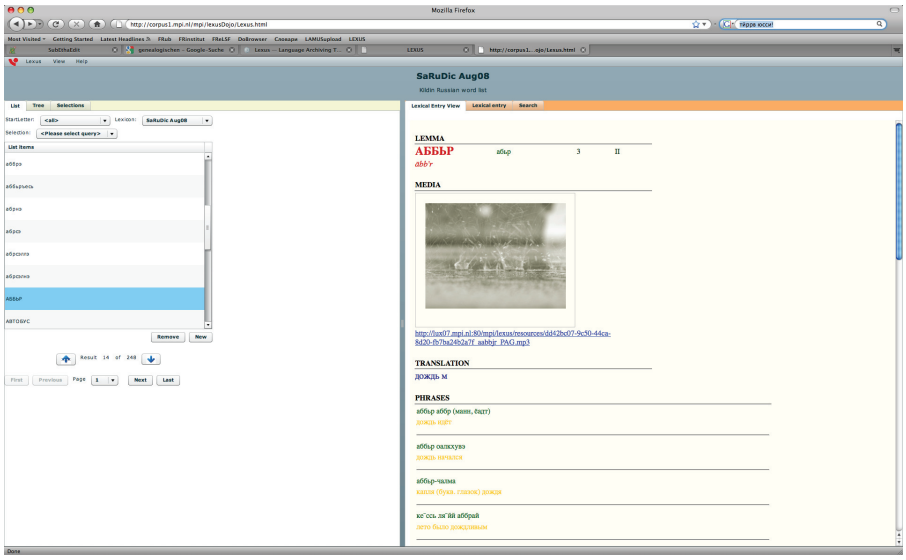

FIGURE 1: Screen shot of Lexical Entry Editor mode

3.2. IMPORTING LEXICAL DATA. Problematic areas for data import from Toolbox to LEXUS are explained in some detail in the short guide "From Toolbox to Lexus" (Ringersma and Rybka 2009b). It is crucial that markers conform to the correct hierarchy in the Toolbox structure file. Toolbox data with a flat structure (e.g., if the Toolbox marker \ge "English gloss" appeared below the marker \lx "lexeme" but not defined as subordinated to the latter) cannot be related to the default components (*Form* and *Sense*) in the LEXUS database, which represents a nested structure by default. Note that it is also recommended that all markers be provided with a descriptive name in the Toolbox structure file, because LEXUS needs the markers' names during import.

XML import was not included as a topic of the manual when this review was written. But importing data from a XML database is generally identical to importing from Toolbox. Since an XML database cannot afford to be "inconsistent" like Toolbox, however, XML import should even be easier.

Once the import function is started from the LEXUS menu in the top frame, a window pops up and the user is guided through the necessary steps in the import procedure. First, you have to browse for the database schema file (e.g. .xsd for a XML database or .typ in Toolbox) and next for the file where the actual data are stored.

3.3. EXPORTING LEXICAL DATA. An export function has yet to be included into the new LEXUS user interface. This is planned for September 2009 (Jacquelijn Ringersma, pers. comm.). Alongside XML and Toolbox formats, export of metadata in IMDI format will be implemented as a function.

3.4. SYNCHRONIZING LEXUS WITH TOOLBOX. Even if LEXUS is chosen as a tool for the creation, presentation, and maintenance of an on-line multimedia lexicon for the language(s) under documentation, many documentary linguists will probably continue with interlinearization and the creation of a preliminary dictionary of morphemes found in the processed texts using Toolbox. One would thus wish for a function to synchronize growing LEXUS and Toolbox projects with each other. For now, however, this task can be performed only with help from the developer team at the MPI Nijmegen. To include the data into the existing lexicon, you have to submit the lexical entries to be added in a Toolbox file.

3.5. ADDING MULTIMEDIA. Any type of file (video, audio, images, and other files like plain text or pdf) can be linked to lexical entries. It is recommended that you create your own data categories for the different kinds of media used in a lexicon. For adding the respective files, the Multimedia tab in the Lexical Entry Editor mode is used.

Media can be uploaded from a local source, or a link can be created to media stored in the MPI archive for linguistic resources. The latter possibility, however, is not yet fully functional.

Media, including links to external sources (e.g., websites or on-line databases like Wikipedia, Google Earth, etc.) or other LEXUS dictionaries, can also be added by defining links through ViCoS.

3.6. SEARCH FUNCTIONS. A search is possible within your own workspace and has two options: a quick search and a query definer. Quick search is available in the Workspace Editor mode or the Lexical Entry Editor mode. This search function works for a simple search in one single lexicon and in one data category. More complex search queries over multiple lexica and in more than one data category can be performed with help of the Query builder. This function even allows you to save queries. It can thus serve as a filter on your wordlist. Complex queries are defined in the Query View tab in the left frame in the Workspace Editor mode. They can be applied in any of the three modes of the program.

3.7. SHARING A LEXUS PROJECT. LEXUS allows sharing lexica with other registered users, either by granting writing rights or only reading rights. Partial writing rights (e.g., providing the right to edit the content of the lexicon, but not its structure) are not possible. Access rights are set in the respective tabs in the Workspace Editor mode.

Several users provided with writing rights can simultaneously edit one and the same LEXUS project in their respective work spaces. Since server-based routines keeping track of different versions are not implemented, changes to the content or the structure of the lexicon may be lost if a second user is saving changes to the same items simultaneously. Even though such a scenario is unlikely to occur, it might be recommended for co-editors of a LEXUS project to agree upon project-internal interaction procedures in order to avoid losing data.

4. LIMITATIONS AND GLITCHES. The new LEXUS user interface shows several improvements when compared to the old version. But it is obvious that the program is still

under development and requirements of divergent groups of users have to be elicited and glitches have to be corrected. As reflected in the missing chapters in the manual, several functions are not yet fully implemented, among them rather important ones such as XML export.[3]

I find the user interface of LEXUS easy to use and the provided functions at the interface more or less self-explanatory. But LEXUS is obviously a tool for the creation and maintenance of a lexicon rather than for the visually appealing presentation of lexical data. Users of LEXUS who are just interested in browsing the data might wish to have an even simpler interface with unnecessary functions hidden, exhibiting more prominent and simpler search functions similar to other on-line dictionaries or databases. A common feature of on-line dictionaries, missing in LEXUS, is a small but centrally placed query window where the word to be looked up can be written. It would probably be a good idea to provide a special interface for simple browsing of the data without requiring a login and password.

Another question connected to the attractiveness of LEXUS for people interested in using (or just browsing) data is the availability of different language localizations. This is a relevant question even for many community-based linguists or other community members who intend to work actively with the tool. Making different language localizations available (e.g., in Spanish, French, Russian, and many other non-English linguae francae) seems imperative in order to introduce LEXUS as a useful tool for practical language work in endangered speech communities.

One rather important shortcoming that I encountered during my work with LEXUS is the lack of Unicode support for input of data other than in standard Roman characters. I came across this glitch when trying to enter Cyrillic characters. This does not work with the pre-installed Cyrillic Unicode keyboards in Mac OS X, nor from the system's Character palette. The same problem also occurs with other non-Roman characters, such as IPA symbols. A preliminary, though unsatisfactory, solution to this problem is to copy-paste the respective words from another application into LEXUS. Note, however, that the problem concerns only Unicode input. The unsupported display of certain Unicode characters, depending on internet browsers' capabilities or the proper configuration of browsers by users, is obviously beyond the sphere of influence of the LEXUS developers.

Another glitch concerns the lexical entry view formatting, which does not always work properly, even on screen resolutions higher than 1280 x 1024 pt. Since externally created HTML layouts give much better results, and are not complicated to create, a few pre-composed templates could be made available directly from the LEXUS website.

5. RECOMMENDATION. LEXUS is definitely a useful tool for language documentation. Being a server-based program that can be accessed by people working at different places makes the tool especially interesting for working groups, e.g., among speech community members or among speech community members and non-local linguists working on and within these communities. Both groups of users are normally interested in making

[3] While this review was being written, new features like applying sort orders and linking resources in the MPI archive were added to LEXUS. Descriptions of these features, however, have yet to be included in the manual.

the collected lexical data available to the public, thus helping to teach and revitalize the language(s) in question. Most linguists, however, would probably also be interested in using the LEXUS data, which will than be corrected and enriched by native collaborators, for further annotation or analyses (lexicographical, etymological, grammatical, etc.). On the other hand, linguists themselves could enrich an ongoing LEXUS lexicon projects with additional data from interlinearizations of newer texts. Synchronizing data between LEXUS and other lexical databases, however, is a problem. In practice, and given the current state of development of LEXUS, a lexicon once started in LEXUS will grow only within itself, i.e., by adding words and explanations by hand, but not by automatic updates from text-analyzing tools, such as the quasi-automatic parser Toolbox.

Another possibility is making LEXUS lexica available in off-line versions, either downloadable from the Internet or distributed on DVD. LEXUS is developed as a server-based on-line tool. But regular export of off-line versions should generally be possible once a practical work flow has been established. Since the XML data of the lexica (including the multimedia) can be exported easily, a LEXUS lexicon should also be transferable to on- or off-line dictionaries, such as the dictionary applications created for different Saami languages by the Giellatekno project.

Primary function:	An on-line dictionary/lexicon editor and viewer
Pros:	Multi-user, on-line, import, export, XML data standard, easy-to-use interface for data input
Cons:	English-only interface (language communities), interface for presentation of data not visually appealing, on-line, program still in early stage of development
Platforms:	Independent (on-line tool)
Open source:	Yes. (A deployment package is not yet available, but the source code can be obtained upon request from Max Planck Institute for Psycholinguistics Nijmegen, The Netherlands)
Proprietary:	n.a. (on-line tool)
Reviewed version:	0.00 (May 2009)
Application size:	n.a. (on-line tool)
Documentation:	User manual (on-line and PDF), website and forum, all found at http://www.lat-mpi.eu/tools/lexus/

References

RIESSLER, MICHAEL, and JACQUELIJN RINGERSMA. 2009. Annotation, archiving and presentation of Kildin Saami lexical data with LEXUS. Ms.

RINGERSMA, JACQUELIJN, and KONRAD RYBKA. 2009a. *Lexus Manual*. Published first version. Downloaded 30 June 2009 from http://www.mpi.nl/corpus/manuals/manual-lexus.pdf.

RINGERSMA, JACQUELIJN, and KONRAD RYBKA. 2009b. *From Toolbox to Lexus: A guide to importing data*. Published first version. Downloaded 30 June 2009 from http://www.mpi.nl/corpus/manuals/toolbox_to_lexus.pdf.

Kristina Kotcheva
kristina.kotcheva@staff.hu-berlin.de